P19.016-Mon

DEEP-STRUCTURE OF THE GENETIC CODE AND THE ORIGIN OF REPLICATION: PATH INVARIANTS AS PRE-LUCA ATTACHMENT SITES

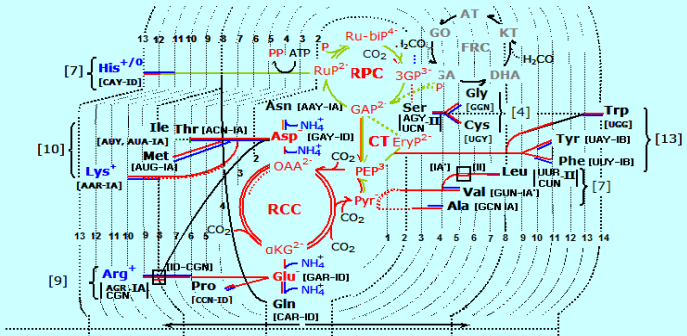
Brian K. Davis:: Research Foundation of Southern California
8861 Villa La Jolla Drive #13595, La Jolla, CA 93037, USA:: davis@resfdnsca.org

Introduction: The genetic code conserves an imprint of its evolution before the Last Universal Common Ancestor (LUCA), over 3.5x10^9 years ago, as its diverse structural regularities are unified by a model equating the time-order of codon assignments to amino acid synthesis path-length. In accord with pre-LUCA attachment to an adaptor/cofactor tRNA, intermediates in the synthesis of all amino acids (except histidine) retain a free α-carboxyl. Attributing the pentose-P cycle invariant PO_4^{2-} to a former poly(P) scaffold, subsequently, led to evidence on the origin of replication.

Method: Pre-RNA replicators, identified by applying the path-invariant principle, were constructed with model-building software (Facio-20.1.3), optimized by restricted Hartree-Fock fragment molecular orbital calculations (Gamess-64-2016).

Results:

□ 5'→3' CODON BASE RECRUITMENT PRODUCED THREE PATH-LENGTH DEPENDENT STAGES IN GENETIC CODE FORMATION



Code Stage	Overprinting	Expansion	N-Fixers	Expansion	Overprinting
Recruited codon base	3'- [2-, 1-sets]	mid- [C→G→U]	5'- [XAN]	mid- [C→G→U]	3'- [2-, 1-sets]
No. stable 4-sets	0	8	0		
Amino acid species	large basic, aromatic	alkyl, hydroxy sulfhydryl	diacid, amides		
No. amino acids	6	10	4		

legend: path-length; scaled path; unscaled path; COO⁻ invariant; NH3+; PO4 2-; RCC reaction cycle; Glu⁻ amino acid [CAR-ID] codon-tRNA type; IA, II tRNA-exchange; [9] no. pre-LUCA enzymes

RCC, reductive citrate cycle. CT, central trunk. RPC, reductive pentose cycle. FRC, formose reaction cycle.

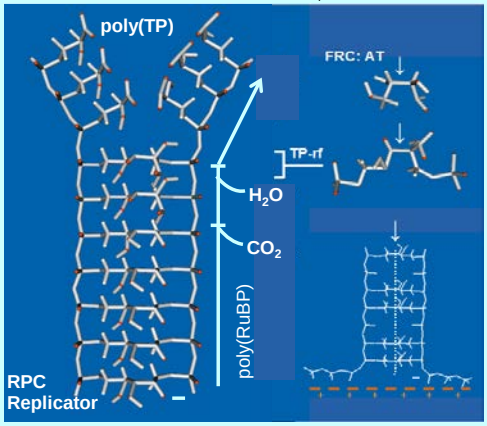
□ PATH-DISTANCE MODEL UNIFIES CODE STRUCTURAL REGULARITIES[1]

Known	New
1. Woese, 1965: NAN, NUN aa hydropathy clusters.	1. Code comprises six domains of contiguous codons read by related pre-LUCA tRNA for same-family aa.
2. Bretscher et al. 1965: nonsense codon inhibition.	2. Amino acid synthesis intermediates retain an invariant α-carboxyl linked to early tRNA-cofactor attachment.
3. Nirenberg et al. 1966: aa biosynthetic clusters.	3. Path-distances reveal codon bases were assigned to distinct kinds of aa in 5'→mid→3' order.
4. Dunnill, 1966: codon 4-sets have 5'/mid-G, C.	4. Compact XAN codon set (X, coding site) first encoded four N-fixer aa (1-2 step paths) and a stop signal.
5. Crick, 1966: universality of standard code.	5. First code places origin of proteins at two RCC N-fixer sites, yielding diacids Asp[1], Glu[1] and amides Asn[2], Gln[2].
6. Wilcox, Nirenberg, 1968: tRNA an aa cofactor.	6. Source duality is amplified in diacid function (aa source v. N donor), phylogenetic depth, and synthetase class.
7. Rodwell, 1969: Ile[2] path has four Val[3] steps.	7. Pre-LUCA tRNA identities indicate Asp[1] was initially precursor to 15 aa, and Glu[1] to only 3 aa.
8. Dillon, 1973: 4-sets predated 2- and 1-sets.	8. Asn[2] and Gln[2] attachment to tRNA blocked lactam formation by these early, labile aa.
9. Dillon, 1973, Wong, 1975: aa coevolved with tRNA.	9. Mid-base was assigned in an (A)→C→G→U order to ten increasingly hydrophobic aa of ~4, 5, and 7 path-steps.
10. Dillon, 1973, Wächtershauser, 1992: aa synthesis paths formed by reductive organo-synthesis.	10. Eight stable code-boxes (4-sets) were assigned (GCN to Om[6]) during expansion from the N-fixer's code.
11. Perlwitz et al. 1988: mid-base most coding capacity.	11. 3'-Base encoded six basic/aromatic aa of 9-14 path-steps, by overprinting six error-prone boxes.
12. Taylor, Coates, 1989: sibling aa share codon 5'-base.	12. tRNA-cofactor exchange led to anomalous assignment of UUR, CUN to Leu[8], and AGR, CGN to Arg[9].
13. Taylor, Coates, 1989: smallest aa assigned 4-sets.	
14. Garrett, Grisham, 1999: na-like aa have long paths.	
15. Lim, Curran, 2001: Y:Y wobble split eight c4-sets.	
16. Brooks et al. 2002: ancient proteins have early aa	
17. Brooks et al. 2002: early aa in ancient proteins.	
18. Brooks, Fresco, 2003: GNN code for early aa.	
19. Biro et al. 2003: codon R, Y mid-base aa clusters.	
20. Norgaard et al. 2009: reconstruction of Pro-Fd-5.	
21. Rodin et al., 2009. tRNA N2:N71 complementarity.	
22. Williams et al. 2009: Synthetase duality.	

Davis, BK 2007. Making sense of the genetic code with the path-distance model. In, *Leading-Edge Messenger RNA Research Communications* Ed., MH Ostrovskiy. New York: Nova Science, Chp. 1, pp. 1-32.
Davis, BK 2013. Making sense of the genetic code with the path-distance model based on tRNA-dependent pathways. https://archive.org/details/MakingSenseOfGeneticCode DOI: 10.13140./RG.2.2.17217.86888

DOI: 10.13140/RG.2.2.20657.68967

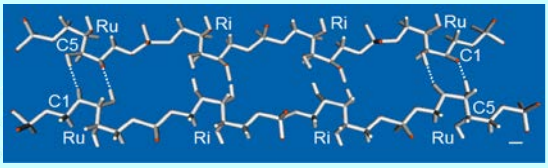
□ RPC REPLICATOR HAS A POLY(PO_4^{2-}) SCAFFOLD



Ladder-like poly(D-ribulose-1,5-bis-phosphate), poly(RuBP), yields poly(triose-P), poly(TP), in an FRC aldotetrose (AT)- like cleavage of the C2-C3 bond. *Lower,* Poly(RuBP) with a complex binary (RuBP, PO_4^{2-}) sequence, within a cationic mineral surface system. rf, replicative-form; bar, 1 Å.

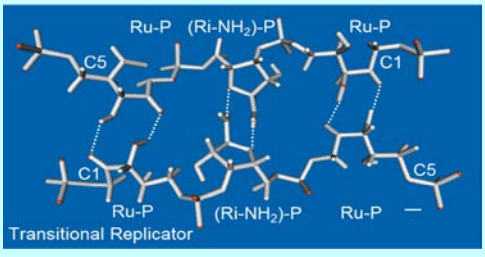
□ ANTI-PARALLEL RuBP MONOMERS H-BOND

□ NON-INTERACTNG Ri-P MONOMERS FORM AN EMBEDDED SCAFFOLD IN A (Ru-P, Ri-P) DUPLEX



Initial duplex ribose pairs, Ri:Ri, are restored on Ru→Ri conversion of unpaired Ru in a poly(Ru-P) daughter strand.

□ NUCLEOBASE-INTERMEDIATES H-BOND, WITH RNA-LIKE EXTERNAL, ANTI-PARALLEL Ri-P SCAFFOLD, IN A PENTOSE-P DUPLEX



Conclusion: Analysis of pre-LUCA reaction sequences furnished evidence on the origin of the genetic code and replication, from spontaneous pre-sugar autocatalysis.